

Combining tumour response and progression free survival as surrogate endpoints for overall survival in advanced colorectal cancer

E.G. Elia^{a,b,*}, N. Stadler^c, O. Ciani^{d,e}, R.S. Taylor^d, S. Bujkiewicz^b

^a Department of Biostatistics, Harvard University, 677 Huntington Ave., Boston, MA 02115, USA

^b Department of Health Sciences, University of Leicester, George Davies Centre, University Road, Leicester LE1 7RH, UK

^c F. Hoffmann-La Roche Ltd, Basel, Switzerland

^d Evidence Synthesis & Modelling for Health Improvement, Institute of Health Research, University of Exeter Medical School, University of Exeter, Exeter EX2 4SG, UK

^e CERGAS Bocconi University, via Rontgen 1, 20136 Milan, Italy

ARTICLE INFO

Keywords:

Surrogate endpoints
Multivariate meta-analysis
Advanced colorectal cancer

ABSTRACT

Background: Progression free survival (PFS) and tumour response (TR) have been investigated as surrogate endpoints for overall survival (OS) in advanced colorectal cancer (aCRC), however their validity has been shown to be suboptimal. In recent years, meta-analytic methods allowing for use of multiple surrogate endpoints jointly have been proposed. Our aim was to assess if PFS and TR used jointly as surrogate endpoints to OS improve their predictive value.

Methods: Data were obtained from a systematic review of randomised controlled trials investigating effectiveness of pharmacological therapies in aCRC, including systemic chemotherapies, anti-epidermal growth factor receptor therapies and anti-angiogenic agents. Multivariate meta-analysis was used to model the association patterns between treatment effects on the surrogate endpoints (TR, PFS) and the final outcome (OS).

Results: Analysis of 33 trials reporting treatment effects on all three outcomes showed reasonably strong association between treatment effects on PFS and OS, however the association parameters were obtained with a large uncertainty. A weak surrogate relationship was noted between the treatment effects on TR and OS. Modelling the two surrogate endpoints, TR and PFS, jointly as predictors of treatment effect on OS gave no marked improvement to surrogate association patterns. Modest improvement in the precision of the predicted treatment effects on the final outcome was noted in studies investigating anti-angiogenic therapy, however it was likely due to chance.

Conclusion: The joint use of two surrogate endpoints did not lead to marked improvement in the association between treatment effects on surrogate and final endpoints in advanced colorectal cancer.

1. Introduction

Surrogate endpoints have been receiving increased attention by the research community in the last three decades as they offer a cost effective and quicker alternative to the use of final outcomes especially if they can be measured with a shorter follow-up period [1]. For surrogate endpoints to be used effectively in clinical research, they need to be validated. There are three levels of surrogate endpoint validation: biological plausibility of association between outcomes, patient-level association between outcomes and study-level association [2,3]. For the purposes of this study we focus on the latter level of validation. Study level association is the hallmark of surrogacy, i.e. establishing whether the treatment effect on the surrogate endpoint is likely to predict a treatment effect on the clinical outcome. This is usually carried out

through meta-analyses of randomised controlled trials (RCTs) and, in particular, using a bivariate meta-analysis [4–6].

To identify a surrogate endpoint for overall survival (OS) in advanced or metastatic colorectal cancer, a number of candidate endpoints have been investigated as potential surrogate endpoints, including progression free survival (PFS), tumour response (TR) or time to progression (TTP) [7–12]. In previous work investigating trial-level surrogacy in advanced colorectal cancer, Buyse et al. found that PFS was an acceptable surrogate endpoint to the overall survival [7]. A more recent study, investigating surrogacy patterns across a broader range of treatments in a meta-analysis of 101 RCTs, showed suboptimal validity of PFS as a surrogate endpoint for OS in advanced or metastatic colorectal cancer [10]. Other studies also investigated the surrogate relationship between the PFS and OS in advanced colorectal cancer that

* Corresponding author at: Department of Biostatistics, Harvard University, 677 Huntington Ave., Boston, MA 02115, USA.

E-mail address: eelia@hsph.harvard.edu (E.G. Elia).

<https://doi.org/10.1016/j.canep.2019.101665>

Received 21 August 2019; Received in revised form 22 November 2019; Accepted 17 December 2019

Available online 05 January 2020

1877-7821/ © 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

suggested that further validation is required [8,9].

Whilst PFS is a common surrogate outcome in clinical oncology, TR has also been often used as a predictor of treatment benefit. Often TR is used as a surrogate marker of treatment effect at the licensing stage of drug development process. For example pertuzumab for the neoadjuvant treatment of HER2-positive breast cancer was approved on the basis of pathological complete response (pCR). In this case, an analysis was undertaken to determine the relationship between pCR and improved event-free and overall survival [13]. Tumour response has been investigated as a potential surrogate endpoint with respect to PFS and OS in several solid or hematologic cancer settings [13–16]. Definition of tumour response is based on objective tumour measurements by imaging methods that allow the classification of patients with a complete or partial confirmed best response as responders. Responses are usually determined according to the Response Evaluation Criteria in Solid Tumors guidelines or the World Health Organization recommendations. As the definition of disease progression to subsequent metastatic stage usually entails clinical (symptoms) and laboratory and imaging-based findings, which also characterise achievement of tumour response, we believe it is reasonable to explore whether the use of treatment effects on two candidate surrogate endpoints, TR and PFS, can result in increased precision of predicted treatment effect on the final outcome, namely OS, when they are modelled jointly.

In these previous studies, using meta-analytic techniques either in the form of a meta-regression or a bivariate meta-analysis, only one surrogate endpoint at a time was investigated. Recently, researchers investigated use of multiple surrogate endpoints in a patient-level surrogate endpoint validation in multiple sclerosis [17] and as joint predictors of clinical benefit measured on final outcome in a meta-analytic framework using Bayesian multi-variate meta-analysis [18]. We provide a rationale for this approach in Section 2.2.

In this paper we investigate whether the use of treatment effects on two candidate surrogate endpoints, TR and PFS, can result in increased precision of predicted treatment effect on the final outcome, namely OS, in advanced colorectal cancer when they are modelled jointly. We investigate the predictive ability of the surrogate endpoints in this setting by conducting a bivariate meta-analysis to investigate one surrogate endpoint at a time and a trivariate meta-analysis to evaluate the surrogate endpoints jointly. We conducted the analysis in the Bayesian framework using multivariate meta-analysis method described by Bujkiewicz et al. [18] as well as introduced here extensions of these methods. The extensions include an alternative parameterisation of the trivariate model and subgroup analysis. The added value of modeling multiple surrogate endpoints jointly was investigated, by comparing the meta-analytic models in terms of surrogacy criteria and predicted intervals.

2. Methods

2.1. Data sources

We used data from a systematic review by Ciani et al. [10], which included treatment effect estimates from 101 randomised controlled trials (RCTs) in advanced or metastatic colorectal cancer that assess pharmacologic therapies against other therapies. The studies included trials investigating a broad range of treatments including five different classes which were systemic chemotherapy, anti-epidermal growth factor receptor monoclonal antibodies (Anti-EGFR), anti-angiogenic agents, other multi-targeted antifolate (MTA) and intra-hepatic arterial chemotherapy (IHA). For the purposes of the systematic review conducted by Ciani et al. [10] the following definitions of outcomes were considered: OS was defined as the time from randomisation to time of death, PFS was defined as the time from randomisation to tumour progression (regardless of how the progression was defined), or death from any cause, and TR was defined by objective tumour measurements by utilising methods that classify patients as responders, with a

complete or partial confirmed best response. Responses were determined using the criteria and recommendations according to the Response Evaluation Criteria in Solid Tumors (RECIST) guidelines [19] or the World Health Organization recommendations [20].

For purpose of this paper, we used data from the above systematic review on treatment effects measured on OS, PFS and TR. Not all of the studies in the systematic review reported treatment effects on all three outcomes of interest. In our main analysis, we used data from trials reporting all three outcomes.

Individual patient data (IPD) were available from one of the RCTs included in the systematic review; study by Hurwitz et al. [21] that investigated the use of bevacizumab in combination with irinotecan, fluorouracil, and leucovorin in patients with metastatic colorectal cancer [21]. The IPD were used to obtain the within-study correlations between the treatment effects on the surrogate endpoints and on the final outcome.

2.2. Statistical analysis

The rationale for modelling the treatment effects on multiple surrogate outcomes simultaneously is that the model can lead to removing some of the measurement error. We can imagine, for example, that three endpoints can be measured sequentially in time: first TR (at year 1) then PFS (year 3) and finally OS (year 5). If these timescales are required to measure treatment effect on these outcomes with reasonable precision, one could imagine that RCTs with, for example, follow up time of 2 years measure precisely only the treatment effect on TR, but there is a lot of uncertainty around the effect on OS and still some uncertainty around the treatment effect on PFS. If we assume that treatment effects on TR are correlated with the effects on PFS and the effects on PFS are correlated with treatment effects on OS, then such model can lead to reduced uncertainty of the treatment effect on PFS using the effect on TR. We cannot measure the true effect on PFS (only have an estimate from a trial), so having measurement on TR can improve estimate of measurement on PFS. This can be useful in particular when the treatment effect on TR is measured precisely and on PFS inaccurately, then the prediction of the true effect on PFS may be more precise (due to accounting for the correlation with the effect on TR) ultimately leading to better prediction of the effect on OS.

We used multivariate meta-analysis in a Bayesian framework to model jointly the treatment effects on one or two surrogate endpoints and on the overall survival which was the final clinical outcome. Bivariate meta-analysis was used to evaluate surrogate endpoint one at the time: TR as a surrogate endpoint to PFS and PFS as a surrogate endpoint to OS, by modelling the treatment effect on these pairs of outcomes. For completeness, we also include the analysis for TR as a surrogate endpoint to OS. Trivariate meta-analysis was used to evaluate both TR and PFS jointly as surrogate endpoints to OS. To model the surrogate endpoints in the sequential order (TR, followed by PFS and then OS as discussed above) we used a model with structured between-studies covariance matrix assuming true treatment effects on TR and OS conditionally independent, but treatment effects on pairs TR-PFS and PFS-OS correlated. The model is discussed in detail in the supplement and the sensitivity analyses to the modelling assumptions was carried out as discussed in Section 2.3.

Treatment effects on PFS and OS were modeled using hazard ratios (HRs) and treatment effects on TR were modelled using odds ratios (ORs). Log scale was used to allow the assumption of normality of the effects. For studies, where there were no responders in one of the treatment arms, continuity correction of 0.5 was added to all values of the contingency table to enable finite odds ratio and variance estimators to be derived.

2.2.1. Surrogacy criteria and cross-validation

We followed the surrogacy criteria introduced by Daniels and Hughes [4], and adopted by Bujkiewicz et al. [18], by which the slope

(and uncertainty around it) indicates the association between the treatment effect on the surrogate endpoint (PFS) and the treatment effect on the final outcome (OS). For the treatment effects to be associated, we require the slope not to be zero. For the association to be perfect the conditional variance (quantifying the degree of variability of the treatment effects around the regression line) would be zero. Moreover, we would expect the intercept to be zero to ensure that no treatment effect on surrogate endpoint will imply no treatment effect on the final outcome. In a similar manner we can describe surrogacy criteria between the first and second surrogate outcomes (TR and PFS), where the second surrogate (here PFS) becomes the final outcome in a bivariate model with a single surrogate endpoint. We also report the adjusted R^2 [25,26] which for perfect surrogate relationship should be one. For further details see Appendix A in the supplementary materials.

In order to investigate whether the joint use of treatment effects on multiple surrogate endpoints gives more precise predictions of the treatment effect on the final outcome, a cross-validation procedure was carried out. In one study at a time the treatment effect on the final outcome was assumed unknown and predicted from the treatment effect on surrogate endpoint (or multiple surrogate endpoints jointly) using the bivariate (or trivariate) meta-analytic model. In this Bayesian approach to multivariate meta-analysis, this was achieved by assuming that the unreported outcomes were missing at random, which were then predicted by the Markov chain Monte Carlo (MCMC) simulation of the model [24]. Further methodological details are included in the supplementary materials. The predicted effects on the final outcome obtained by the bivariate and the trivariate meta-analytic models were compared in terms of the width of the predicted intervals. We investigated surrogacy across all RCTs as well as in subgroups of class of therapy.

2.3. Sensitivity analyses

We carried out sensitivity analysis to the modelling assumptions, but carrying out the analysis using an alternative model with a structured between-studies covariance matrix (assuming this time the treatment effects on TR and PFS conditionally independent but both correlated with the treatment effect on OS) and a model with an unstructured between-studies covariance matrix (assuming that the true treatment effects on all three outcomes were correlated). The models are described in details in the online appendix.

We carried out a sensitivity analysis on a larger data set including studies which reported treatment effects on at least two outcomes. Analyses were also repeated in subsets of data defined by the class of treatment. A sensitivity analyses to examine the impact of an outlying observation (with large treatment effect on TR, which can be observed in Fig. 1) and the choice of the prior distribution for the between-studies correlation were also carried out (for details see supplementary materials, Appendix A).

Crossover in RCTs, for example from the control to experimental arm following progression, often results in loss of information about the treatment effect on the final outcome; what the effect would have been if crossover was not allowed. As patients move to the experimental treatment arm, the difference in treatment effect on OS between the treatment arms diminishes, leading potentially to zero effect with large uncertainty. This creates difficulty in estimating the association patterns between treatment effects on surrogate and final endpoint and the estimates of the latter are not reliable and their variability is reduced (potentially diminishing the correlation between the treatment effect on the two outcomes). Another sensitivity analysis was carried out on the subset of trials which did not allow for crossover.

2.4. Software and computing

All models were implemented in WinBUGS [27] where the estimates were obtained using MCMC simulation using 250,000 iterations

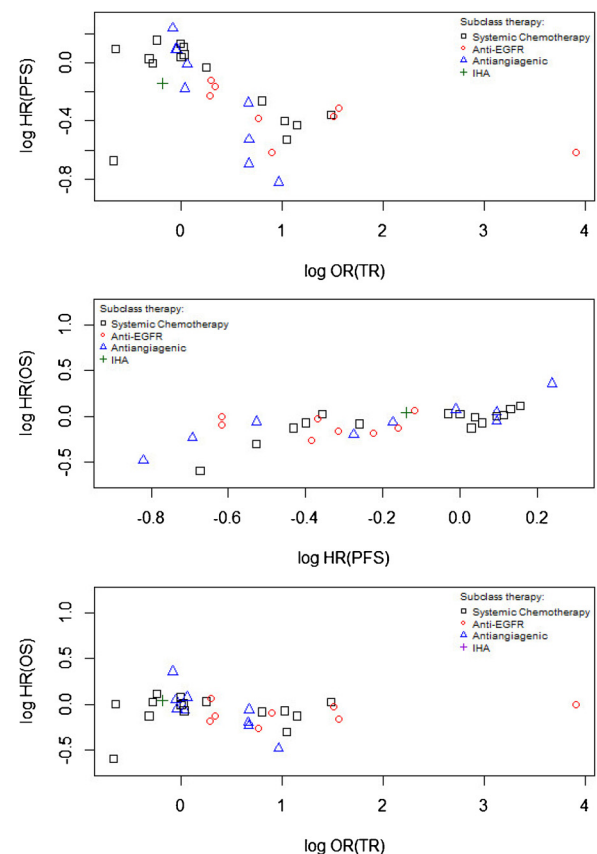


Fig. 1. Scatter plot of treatment effect estimates on PFS vs TR (top), OS vs PFS (middle) and OS vs TR (bottom); for each subclass of therapy for the main data set including trials reporting all three outcomes.

(including 150,000 burn-in). Convergence was checked by visually assessing the history, chains and autocorrelation using graphical tools in WinBUGS. All posterior estimates are presented as means with the 95% credible intervals (CrI). R was used for data manipulation and to execute WinBUGS code multiple times (for validation of surrogates for each study) using the R2WinBUGS package [28]. OpenBUGS and R2OpenBUGS version of the software was used for the cross-validation procedures which were conducted using Linux (Red Hat, Inc., Raleigh, North Carolina)-based high performance computer. WinBUGS programs corresponding to the bivariate and trivariate models are included in the online supplementary materials, Appendix B.

3. Results

3.1. Summary of the data

Out of the 101 RCT's 99 reported estimates of treatment effects on at least one of the three outcomes of interest, 51 studies reported the treatment effects on at least two of the outcomes and 33 studies reported treatment effects on all three outcomes. In the main analysis of data from studies reporting all three outcomes (33 studies), subgroups of therapy included 15 studies which investigated the use of chemotherapy, eight studies which evaluated the use of anti-EGFR therapies, nine studies investigating the use of anti-angiogenic agents and one study evaluating the use of IHA. More details about the study characteristics are included in Ciani et al. [10].

In the sensitivity analysis of studies reporting at least two outcomes, we used 51 of which: 48 studies reported treatment effect on OS (23 investigating chemotherapy, 11 anti-EGFR, 12 anti-angiogenic agents, 1 MTA, 1 IHA), 39 studies reported treatment effect on PFS (17 trials investigating chemotherapy, 9 anti-EGFR, 12 anti-angiogenic agents,

and 1 IHA), and 48 studies reported treatment effect on TR (25 trials of chemotherapy, 10 anti-EGFR, 11 anti-angiogenic agents, 1 MTA, 1 IHA). The results obtained from the trivariate meta-analysis, where at least two outcomes were reported (with some missing data), were compared to the bivariate analysis of the complete data for each surrogacy pair: 45 studies reported treatment effects on both TR and OS (23 trials of chemotherapy, 10 anti-EGFR, 10 anti-angiogenic agents, 1 MTA and 1 IHA); 36 studies reported treatment effects on TR and PFS (17 trials of chemotherapy, 8 anti-EGFR, 10 anti-angiogenic agents, 1 IHA); 36 studies reported treatment effects on PFS and OS (15 trials of chemotherapy, 9 anti-EGFR, 11 anti-angiogenic agents and 1 IHA). List of references for studies included in the analysis can be found in the online supplementary materials (Appendix C: reference lists A and B).

In the sensitivity analysis of trials that did not allow for patient crossover we combined seven studies, out of the total 33 trials reporting all three outcomes. The studies included three trials of chemotherapy, one anti-EGFR therapy, two anti-angiogenic agents and one IHA.

Data for the main analysis are presented in Fig. 1 in the form of scatter plots of the treatment effects on each pair of outcomes (PFS vs TR, OS vs PFS and OS vs TR). The plots show a possible strong positive association between the treatment effects on PFS and OS and negative association for the treatment effects on TR and PFS (increased response rate is expected to lead to reduced progression and hence a negative correlation between the treatment effects on these outcomes), but likely weak negative or no association between the treatment effects on TR and OS.

Exploratory analysis of the data showed a lot of heterogeneity of the treatment effects for TR and PFS, with the confidence intervals of the treatment effects on TR particularly wide, especially for two classes of therapy: the chemotherapy and anti-EGFR therapies. Further details, along with the estimates of the within-study correlations, are included in the supplementary materials (Appendix D and Table 1 of Appendix E).

3.2. Results of main analysis

Table 1 shows results of three bivariate meta-analyses and the trivariate meta-analysis (discussed in the first part of Section 2.2) conducted using the data from studies reporting treatment effects on all of the three outcomes. The table lists the surrogacy criteria for all surrogate relationships (and both the bivariate and trivariate models). The top part of the table shows results of applying the models to all of the data. They are followed by the results of three sets of analyses of applying the models to subsets of the data defined by the class of therapy.

Results of the three bivariate models applied to all of the data showed that there was an association between the treatment effects on each pair of outcomes. The intervals of the intercepts obtained from the bivariate models all contained zero indicating that no effect on the surrogate endpoint could imply no effect on the final outcome. The intervals for the slopes did not contain zero indicating positive association where slope was positive and negative association where slope was negative. However the surrogate relationships were not strong. When investigating TR as a surrogate endpoint for OS, the association between the treatment effects on the two outcomes was weak in terms of the small slope $\lambda_{21} = -0.05$ (95% CrI: $-0.13, 0.00$) and the 95% CrI contained zero when rounded to the second decimal place, and the mean and the lower bound of $R^2_{adj} = 0.33$ (95% CrI: $0.00, 0.91$) were also small. The slope and the adjusted R-squared were higher for the relationship between the treatment effects on TR and PFS; slope was -0.32 (95% CrI: $-0.45, -0.20$) and $R^2_{adj} = 0.61$ (95% CrI: $0.27, 0.87$). However the lower bound of the CrI for the R-squared was still low and the conditional variance was relatively high, 0.02 (95% CrI: $0.01, 0.05$), indicating a weak surrogate relationship. The surrogate relationship between the treatment effects on PFS and OS appeared stronger in terms of the conditional variance 0.00 (95% CrI: $0.00, 0.01$) and the $R^2_{adj} = 0.58$ (95% CrI: $0.06, 0.97$) with lower ends of CrIs being close to zero.

Results from the trivariate meta-analysis, which described the associations between the treatment effects on the two pairs of outcomes (effects on TR and PFS and effects on PFS and OS) in a single model,

Table 1

Surrogacy criteria obtained from three bivariate models and a trivariate model for the association between treatment effects on the surrogate (TR or PFS) and the final outcome (OS or PFS in one of the bivariate analyses). The results are posterior means and 95% credible intervals.

	Bivariate analyses			Trivariate analysis	
	TR OS	TR PFS	PFS OS	TR PFS	PFS OS
<i>All treatments (N = 33)</i>					
Intercept	$-0.03(-0.07, 0.02)$	$-0.05(-0.14, 0.02)$	$-0.02(-0.06, 0.03)$	$-0.05(-0.13, 0.02)$	$-0.02(-0.06, 0.03)$
Slope	$-0.05(-0.13, 0)$	$-0.32(-0.45, -0.2)$	$0.22(0.03, 0.41)$	$-0.31(-0.43, -0.19)$	$0.19(0.02, 0.4)$
Variance	$0(0, 0.01)$	$0.02(0.01, 0.05)$	$0(0, 0.01)$	$0.02(0.01, 0.04)$	$0(0, 0.01)$
$R^2_{adjusted}$	$0.33(0, 0.91)$	$0.61(0.27, 0.87)$	$0.58(0.06, 0.97)$	$0.64(0.3, 0.89)$	$0.5(0.02, 0.95)$
<i>Systemic chemotherapy (N = 15)</i>					
Intercept	$-0.02(-0.08, 0.04)$	$-0.04(-0.16, 0.06)$	$-0.02(-0.08, 0.04)$	$-0.04(-0.14, 0.05)$	$-0.02(-0.08, 0.04)$
Slope	$-0.03(-0.11, 0)$	$-0.26(-0.42, -0.08)$	$0.17(0, 0.45)$	$-0.25(-0.4, -0.09)$	$0.14(0, 0.4)$
Variance	$0(0, 0.01)$	$0.02(0, 0.08)$	$0(0, 0.01)$	$0.02(0, 0.06)$	$0(0, 0.01)$
$R^2_{adjusted}$	$0.39(0, 0.96)$	$0.58(0.07, 0.96)$	$0.52(0.01, 0.98)$	$0.66(0.11, 0.98)$	$0.47(0, 0.97)$
<i>Anti-EGFR therapies (N = 8)</i>					
Intercept	$-0.06(-0.16, 0.09)$	$-0.21(-0.37, 0.01)$	$-0.06(-0.16, 0.14)$	$-0.18(-0.37, 0.05)$	$-0.04(-0.16, 0.19)$
Slope	$-0.04(-0.18, 0)$	$-0.14(-0.36, -0.01)$	$0.14(0, 0.63)$	$-0.16(-0.39, -0.02)$	$0.17(0, 0.78)$
Variance	$0.01(0, 0.05)$	$0.02(0, 0.08)$	$0.01(0, 0.04)$	$0.02(0, 0.1)$	$0.01(0, 0.05)$
$R^2_{adjusted}$	$0.19(0, 0.82)$	$0.45(0, 0.96)$	$0.22(0, 0.85)$	$0.5(0.01, 0.97)$	$0.19(0, 0.83)$
<i>Anti-angiogenic agents (N = 9)</i>					
Intercept	$0.04(-0.09, 0.2)$	$0.03(-0.18, 0.25)$	$0.02(-0.1, 0.15)$	$0.03(-0.18, 0.25)$	$0.02(-0.09, 0.14)$
Slope	$-0.35(-0.89, -0.03)$	$-0.87(-1.64, -0.3)$	$0.38(0.05, 0.79)$	$-0.85(-1.64, -0.27)$	$0.37(0.04, 0.77)$
Variance	$0.02(0, 0.07)$	$0.04(0, 0.15)$	$0.02(0, 0.06)$	$0.03(0, 0.14)$	$0.01(0, 0.06)$
$R^2_{adjusted}$	$0.52(0.01, 0.97)$	$0.74(0.13, 0.99)$	$0.59(0.03, 0.97)$	$0.72(0.1, 0.99)$	$0.56(0.02, 0.97)$

TR – tumour response, PFS – progression free survival, OS – overall survival.

were similar to those obtained from separate bivariate models. Precision around the intercept, slope and the conditional variance was minimally reduced for the association between the treatment effects on TR and PFS in the trivariate analysis whereas precision for these estimates for the association between the treatment effects on PFS and OS remained largely the same in the trivariate analysis as in the bivariate analysis using PFS as a single surrogate endpoint.

Table 1 also shows the results for each subclass of therapy. For subgroup of trials investigating systemic chemotherapy, the results were similar to those obtained from the whole cohort of studies but typically obtained with increased uncertainty (wider CrIs) and weaker association in terms of lower mean slope. The adjusted R-squared was minimally higher in this subgroup for the association between the treatment effects on TR and OS, whilst a lower mean slope was obtained for the association between TR and PFS and between PFS and OS, compared to the analysis of all treatments. The slopes for the association between the treatment effects on TR and PFS and between the effects on PFS and OS were obtained with minimally higher precision when using the trivariate model compared to the bivariate analysis. For the anti-EGFR therapies, also similar results were obtained to those from the analysis conducted on data from all of the studies but also, similarly as for systemic chemotherapy, with weaker association pattern. For anti-angiogenic agents the mean slopes and the mean R-squared values were considerably higher for all investigated surrogacy relationships compared to other subclasses and the analysis of all treatments, however they were obtained with high uncertainty, also likely due to the small number of studies in the subgroup.

3.2.1. Cross-validation

To investigate the predictive value of the surrogate endpoints when modelled jointly, a cross-validation procedures were carried out. The predicted treatment effects on OS predicted from the treatment effect on PFS alone were compared with those predicted from treatment effects on both TR and PFS jointly by exploring the associated uncertainty described by the predicted intervals as well as whether the predicted intervals contained the observed point estimate of the treatment effect on OS in each study. When looking at predicted treatment effects obtained using the complete data set of the 33 studies, some of the predicted intervals were inflated when making predictions from the treatment effect on both surrogate endpoints jointly, compared to the predictions made from the treatment effect on PFS only. The intervals were on average 0.21% wider with the percentage change of the width of the interval ranging between 1.96% reduction to 4.4% increase. However, from the point of view of the cross-validation procedure, the 95% predicted intervals included the observed point estimate in most of the studies apart from one which was an extreme lowest value of the treatment effect on OS. This was the case for both bivariate and trivariate models. Full set of predicted intervals is included in the supplementary materials (Table 2 in Appendix E).

The cross-validation procedure was repeated for each class of therapy separately. In terms of predicted intervals, on average there was a very modest improvement in precision of predictions for the anti-

angiogenic agents obtained from the trivariate compared to the bivariate model (on average 2.35% narrower 95% predicted interval). For the systemic chemotherapy class there was a minimal improvement in precision for most of the studies, with an average percentage reduction of 1.41% predicted interval. However, the predicted intervals were inflated for the anti-EGFR therapies with, on average, 6.9% increase in the width of the interval. All predicted intervals are included in the supplementary materials (Table 3 in Appendix E).

3.2.2. Summary of main results

Overall there was not much benefit of combining treatment effects on two surrogate endpoints to predict the treatment effect on the final outcome. This lack of improvement, or even increased uncertainty of the predicted effect when using multiple surrogate endpoints, may be due to increased overall between-studies heterogeneity when extending the data to include the treatment effect on TR. The between-studies heterogeneity for the treatment effect on TR was considerably higher compared to the heterogeneity of the treatment effects on PFS and OS in the data set including all treatments. This was also the case for the subgroups of studies including the systemic chemotherapy and the anti-EGFR therapy trials. However, for the anti-angiogenic agents, the between-studies heterogeneity of the treatment effect on TR was comparable with that for the treatment effects on PFS. This may explain some increase in precision of the slope and the predicted effects on OS when using multiple surrogate endpoints in this class of therapy, as including additional outcome did not increase overall uncertainty. However, due to small number of studies the added value was minimal. The treatment effects on all three outcomes are comparable between those from bivariate and trivariate models. All mean treatment effects and the heterogeneity parameters are listed in the supplementary materials (Table 1 in Appendix E).

3.3. Results of sensitivity analyses

Two alternative parameterisations of the trivariate model (described in Section 2.3) were investigated to assess their impact on the results. For the first alternative structured model the parameters describing the association pattern between the treatment effects on PFS and OS are similar to those obtained from the main trivariate model. For the two structured models the slope is $\lambda_{32} = 0.19$ (95% CrI: 0.02, 0.4) and 0.13 (95% CrI: 0.01, 0.33) respectively. The adjusted R-squared for the alternative structured model was reduced but was obtained with higher precision compared to the main structured model; R_{adj}^2 becomes 0.35 (95% CrI: 0, 0.87) compared to 0.5 (95% CrI: 0.02, 0.95) in the main model. For the unstructured trivariate model the treatment effects on the two outcomes showed higher association in terms of slope $\lambda_{32} = 0.29$ (95% CrI: -0.01, 0.73) but the association was found to be only marginal (95% CrI contained zero) (Table 2).

Sensitivity analysis, extending the data set to the 51 studies reporting at least two outcomes gave similar results to the main analysis; surrogacy criteria for the association between the treatment effects on PFS and OS were satisfied both when looking at all therapies as well as

Table 2

Surrogacy criteria obtained from the bivariate model and three trivariate models: the main structured model (assuming conditional independence between true effects on TR and OS), the alternative structured model (assuming conditional independence between true effects on TR and PFS) and the unstructured model (all true effect on the three outcomes are correlated), applied to the complete data set (33 studies).

	PFS OS bivariate model	Trivariate models		
		PFS OS structured main model	PFS OS structured alternative model	PFS OS unstructured model
Intercept	-0.02(-0.06, 0.03)	-0.02(-0.06, 0.03)	-0.02(-0.06, 0.03)	-0.06(-0.14, 0.01)
Slope	0.22(0.03, 0.41)	0.19(0.02, 0.4)	0.13(0.01, 0.33)	0.29(-0.01, 0.73)
Variance	0(0, 0.01)	0(0, 0.01)	0(0, 0.01)	0(0, 0.01)
$R_{adjusted}^2$	0.58(0.06, 0.97)	0.5(0.02, 0.95)	0.35(0, 0.87)	0.37(0.04, 0.89)
DIC	-116.92(-136.5, -95.37)	-117.61(-144.9, -87.1)	-119.96(-146.6, -91.58)	-120.77(-148, 90.75)

for chemotherapy and anti-angiogenic agents (Tables 4 and 5 in Appendix E in the online supplementary materials).

Results of the sensitivity analysis of trials with no crossover, presented in Appendix E (Section 5.3) of the online supplementary materials, were similar to those obtained from the main analysis with respect to the surrogacy criteria. The treatment effects on the surrogate and final outcomes appeared to be associated for all investigated surrogacy relationships. However for all parameters describing surrogate relationships for both the bivariate and the trivariate analyses were obtained with larger uncertainty compared to the whole set of studies, which was most likely due to the small number of studies (only 7) in the meta-analysis. The association between treatment effects on PFS and OS appeared to be weaker in terms of the conditional variance which increased compared to the results obtained from all 33 studies reporting all three outcomes. This was the case in both sets of results, from the bivariate and the trivariate analyses. This also led to reduced values of the adjusted *R*-squared. The use of multiple surrogate endpoints did not improve the strength of the association patterns when compared to the use of a single surrogate endpoint in this subset of studies. Similarly as for the full data set, this could also be due to the large heterogeneity of the treatment effects on TR including the average effects and the heterogeneity parameters. The cross validation showed on average a very modest increase in precision, by on average 2.3% and up to 6.2%, when predicting the treatment effect on OS from the treatment effects on both PFS and TR compared to the predictions made from the effect on PFS alone. Full set of results from the cross validation are included in Table 8 of the online supplementary materials.

An additional sensitivity analysis was carried out to investigate an impact of an outlying observation (study with the largest effect size estimate for TR). The results (listed in Table 9 of the online supplementary materials) showed an increased mean slope and the mean *R*-squared for the full set of studies and the subgroups of the EGFR inhibitors, however the uncertainty around these parameters in both sets of results also increased.

A final sensitivity analysis was carried out investigating the impact of the choice of the prior distribution for the between-studies correlation, replacing the informative prior distributions $\text{uniform}(-1,0)$ for the negative association and $\text{uniform}(0,1)$ for the positive association with a non-informative prior distributions $\text{uniform}(-1,1)$. The results of this analysis (listed in Table 10 of the online supplementary materials) were somewhat different compared to those obtained from the main analysis, in particular for anti EGFR therapies where the slope for the association between the treatment effects on PFS and OS become negative with large uncertainty, confirming that this prior was not suitably vague as discussed by Burke et al. [22].

4. Discussion

We investigated the use of multiple surrogate endpoints as joint predictors of the clinical benefit measured on the final clinical outcome in advanced colorectal cancer. A multivariate meta-analytic framework allowed us to combine treatment effects on two candidate surrogate endpoints (TR and PFS) and the treatment effect on the final clinical outcome (OS). In a Bayesian meta-analytic framework, we modelled the correlated treatment effect in the product normal formulation which is a convenient form to explore a range of parameters describing the surrogacy relationships, such as the intercept, slope, conditional variance (as set out by Daniels and Hughes [4]) and the adjusted *R*-squared (introduced by Burzykowski et al. [26] and in the Bayesian framework by Renfro et al. [25]). These models also are used to make predictions of the treatment effect on the final clinical outcome (OS) from the treatment effect on the surrogate endpoints. In this respect they have an advantage of taking into account of the uncertainty around all the parameters, including the measurement error around the treatment effects on surrogate endpoints (in contrast to, for example, the standard approach to meta-regression where treatment effects on surrogate

endpoints are treated as fixed covariates) [18].

The treatment effects on PFS and TR were associated with treatment effect on OS. However, overall the joint use of two surrogate endpoints did not lead to much improvement in the association between treatment effects on the surrogate and final endpoints but in the subclass of anti-angiogenic agents led to very modest improvement in precision of the predicted effects on OS. Some small improvement in precision, when modelling both surrogate endpoints jointly, was also observed in cross-validation procedure conducted on trials without cross-over. In the trials allowing for cross-over, there is typically reduced effect on OS with large uncertainty around the treatment effect estimate. This is likely to affect the results of modelling surrogate relationships, using both the bivariate and the trivariate methods. It is possible that the trivariate approach would show some noticeable benefit if more studies were available in the analysis of studies without the cross-over. In our analysis of the trials which did not allow for the cross-over, there was typically reduced uncertainty of the predicted effects when using multiple surrogates, but the reduction was small as the number of studies in the analysis was also small, therefore we cannot draw strong conclusions based on our findings. Not all studies reported whether the treatment cross-over was allowed. Another source of uncertainty that may have prevented the improvement of the surrogate relationship when using both candidate surrogate endpoints was the large between-studies heterogeneity of the treatment effect on the tumour response. This may have been caused by the heterogeneity of the methods used to measure the response across the trials.

A limitation of the data was also availability of IPD from only one RCT. Therefore we assumed that the within-study correlations were the same across studies. This is, however, not an unusual approach and this assumption has been made by other authors [29,30]. This assumption should not have substantial impact on the results, unless the actual within-study correlations vary substantially across the studies. However, a recent study by Papanikos et al. [31], looked into within-study correlations between the treatment effects on the same outcomes from several RCTs in advanced colorectal cancer and they were all very similar across these studies, suggesting suitability of this assumption for our data.

In conclusion, the impact of the joint modelling of the treatment effects on two surrogate endpoints (TR and PFS), on their surrogate relationship with the treatment effect on the overall survival was not noticeable in advance colorectal cancer. Further work will be needed to investigate in detail the treatment cross-over and the heterogeneity of the definitions of the outcomes and potentially the patient populations.

Authors' contribution

SB conceived the research idea and revised the manuscript to improve general readability. EE prepared the data for analysis and undertook the analyses and interpretation under the supervision and feedback from SB. EE drafted the paper and revised it following comments from all of the authors. NS contributed to data analysis and interpretation of the results and critically revised the manuscript. OC and RT contributed data from their systematic review and critically revised the manuscript. All authors approved the final version of the manuscript.

Conflict of interest

Nicolas Stadler is employed by F. Hoffman-La Roche. Sylwia Bujkiewicz has served as a speaker and a consultant for Roche.

Acknowledgements

This work was supported by the Medical Research Council (MRC) Methodology Research Programme [New Investigator Research Grant MR/L009854/1 awarded to Sylwia Bujkiewicz]. Eleni Elia and Sylwia

Bujkiewicz were financially supported by this grant while being employed by the University of Leicester. The funder had no involvement in the conduct of this research. The research presented here used the ALICE High Performance Computing Facility at the University of Leicester.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.canep.2019.101665>.

References

- [1] T. Burzykowski, G. Molenberghs, M. Buyse, *The Evaluation of Surrogate Endpoints*, Springer Science & Business Media, 2006.
- [2] R.S. Taylor, J. Elston, The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK health technology assessment reports, *Health Technol. Assess.* 13 (8) (2009).
- [3] O. Ciani, M. Buyse, M. Drummond, G. Rasi, E.D. Saad, R.S. Taylor, Time to review the role of surrogate end points in health policy: state of the art and the way forward, *Value Health* 20 (3) (2017) 487–495.
- [4] M.J. Daniels, M.D. Hughes, Meta-analysis for the evaluation of potential surrogate markers, *Stat. Med.* 16 (17) (1997) 1965–1982.
- [5] M. Buyse, G. Molenberghs, T. Burzykowski, D. Renard, H. Geys, The validation of surrogate endpoints in meta-analyses of randomized experiments, *Biostatistics* 1 (1) (2000) 49–67.
- [6] S. Bujkiewicz, J.R. Thompson, E. Spata, K.R. Abrams, Uncertainty in the Bayesian meta-analysis of normally distributed surrogate endpoints, *Stat. Methods Med. Res.* 26 (5) (2017) 2287–2318.
- [7] M. Buyse, T. Burzykowski, K. Carroll, S. Michiels, D.J. Sargent, L.L. Miller, G.L. Elfring, J.-P. Pignon, P. Piedbois, Progression-free survival is a surrogate for survival in advanced colorectal cancer, *J. Clin. Oncol.* 25 (33) (2007) 5218–5224.
- [8] C. Giessen, R.P. Laubender, D.P. Ankerst, S. Stintzing, D.P. Modest, U. Mansmann, V. Heinemann, Progression-free survival as a surrogate endpoint for median overall survival in metastatic colorectal cancer: literature-based analysis from 50 randomized first-line trials, *Clin. Cancer Res.* 19 (1) (2013) 225–235.
- [9] Q. Shi, A. De Gramont, A. Grothey, J. Zalcberg, B. Chibaudel, H.-J. Schmoll, M.T. Seymour, R. Adams, L. Saltz, R.M. Goldberg, et al., Individual patient data analysis of progression-free survival versus overall survival as a first-line end point for metastatic colorectal cancer in modern randomized trials: findings from the analysis and research in cancers of the digestive system database, *J. Clin. Oncol.* 33 (1) (2015) 22–28.
- [10] O. Ciani, M. Buyse, R. Garside, J. Peters, E.D. Saad, K. Stein, R.S. Taylor, Meta-analyses of randomized controlled trials show suboptimal validity of surrogate outcomes for overall survival in advanced colorectal cancer, *J. Clin. Epidemiol.* 68 (7) (2015) 833–842.
- [11] C. Chirila, D. Odom, G. Devercelli, S. Khan, B.N. Sherif, J.A. Kaye, I. Molnár, B. Sherrill, Meta-analysis of the association between progression-free survival and overall survival in metastatic colorectal cancer, *Int. J. Colorectal Dis.* 27 (5) (2012) 623–634.
- [12] P.A. Tang, S.M. Bentzen, E.X. Chen, L.L. Siu, Surrogate end points for median overall survival in metastatic colorectal cancer: literature-based analysis from 39 randomized controlled trials of first-line chemotherapy, *J. Clin. Oncol.* 25 (29) (2007) 4562–4568.
- [13] P. Cortazar, L. Zhang, M. Untch, K. Mehta, J.P. Costantino, N. Wolmark, H. Bonnefoi, D. Cameron, L. Gianni, P. Valagussa, S.M. Swain, Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis, *Lancet* 384 (2014) 164–172.
- [14] G. Nakayama, T. Fujii, K. Murotani, K. Uehara, N. Hattori, M. Hayashi, C. Tanaka, D. Kobayashi, M. Kanda, S. Yamada, H. Sugimoto, Modified two-dimensional response as surrogate marker of overall survival in patients with metastatic colorectal cancer, *Cancer Sci.* 107 (2016) 1492–1498.
- [15] G. Roviello, F. Andre, S. Venturini, B. Pistilli, G. Curigliano, M. Cristofanilli, P. Rosellini, D. Generali, Response rate as a potential surrogate for survival and efficacy in patients treated with novel immune checkpoint inhibitors: a meta-regression of randomised prospective studies, *Eur. J. Cancer* 86 (2017) 257–265.
- [16] T. Burzykowski, M. Buyse, M.J. Piccart-Gebhart, G. Sledge, J. Carmichael, H.J. Lck, J.R. Mackey, J.M. Nabholz, R. Paridaens, L. Biganzoli, J. Jassem, Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer, *J. Clin. Oncol.* 26 (12) (2008) 1987–1992.
- [17] M.P. Sormani, L. Bonzano, L. Roccatagliata, G.R. Cutter, G.L. Mancardi, P. Bruzzi, Magnetic resonance imaging as a potential surrogate for relapses in multiple sclerosis: a meta-analytic approach, *Ann. Neurol.* 65 (3) (2009) 268–275.
- [18] S. Bujkiewicz, J.R. Thompson, R.D. Riley, K.R. Abrams, Bayesian meta-analytical methods to incorporate multiple surrogate endpoints in drug development process, *Stat. Med.* 35 (7) (2016) 1063–1089, <https://doi.org/10.1002/sim.6776>.
- [19] P. Therasse, S.G. Arbuck, E.A. Eisenhauer, J. Wanders, R.S. Kaplan, L. Rubinstein, J. Verweij, M. Van Glabbeke, A.T. van Oosterom, M.C. Christian, et al., New guidelines to evaluate the response to treatment in solid tumors, *J. Natl. Cancer Inst.* 92 (3) (2000) 205–216.
- [20] W.H. Organization, et al., *Who Handbook for Reporting Results of Cancer Treatment*, (1979).
- [21] H. Hurwitz, L. Fehrenbacher, W. Novotny, T. Cartwright, J. Hainsworth, W. Heim, J. Berlin, A. Baron, S. Griffing, E. Holmgren, et al., Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer, *N. Engl. J. Med.* 350 (23) (2004) 2335–2342.
- [22] D.L. Burke, S. Bujkiewicz, R.D. Riley, Bayesian bivariate meta-analysis of correlated effects: impact of the prior distributions on the between-study correlation, borrowing of strength, and joint inferences, *Stat. Methods Med. Res.* 27 (2) (2018) 428–450.
- [24] S. Bujkiewicz, J.R. Thompson, A.J. Sutton, N.J. Cooper, M.J. Harrison, D.P. Symmons, K.R. Abrams, Multivariate meta-analysis of mixed outcomes: a Bayesian approach, *Stat. Med.* 32 (22) (2013) 3926–3943.
- [25] L.A. Renfro, Q. Shi, D.J. Sargent, B.P. Carlin, Bayesian adjusted r^2 for the meta-analytic evaluation of surrogate time-to-event endpoints in clinical trials, *Stat. Med.* 31 (8) (2012) 743–761.
- [26] T. Burzykowski, G. Molenberghs, M. Buyse, H. Geys, D. Renard, Validation of surrogate end points in multiple randomized clinical trials with failure time end points, *J. R. Stat. Soc.: Ser. C Appl. Stat.* 50 (4) (2001) 405–422.
- [27] D.J. Lunn, A. Thomas, N. Best, D. Spiegelhalter, Winbugs – a Bayesian modelling framework: concepts, structure, and extensibility, *Stat. Comput.* 10 (4) (2000) 325–337, <https://doi.org/10.1023/A:1008929526011>.
- [28] S. Sturtz, U. Ligges, A. Gelman, R2winbugs: a package for running winbugs from R, *J. Stat. Softw.* 12 (3) (2005) 1–16 <http://www.jstatsoft.org>.
- [29] Y. Wei, J.P.T. Higgins, Estimating within-study covariances in multivariate meta-analysis with multiple outcomes, *Stat. Med.* 32 (2013) 1191–1205.
- [30] S. Bujkiewicz, D. Jackson, J.R. Thompson, R.M. Turner, N. Stadler, K.R. Abrams, I.R. White, Bivariate network meta-analysis for surrogate endpoint evaluation, *Stat. Med.* 38 (2019) 3322–3341.
- [31] T. Papanikos, J.R. Thompson, K.R. Abrams, N. Stadler, O. Ciani, R. Taylor, S. Bujkiewicz, A Bayesian Hierarchical Meta-Analytic Method for Modelling Surrogate Relationships That Vary Across Treatment Classes, (2019) [arXiv:1905.07194](https://arxiv.org/abs/1905.07194).